Les risques majeurs d'Ipv6 pour la protection des données - Le déploiement du nouveau plan de numérotation du réseau Internet et ses risques majeurs pour la protection des données

Dinant, Jean-Marc

Published in:
La 23ème conférence internationale des commissaires à la protection des données vie privée - droit de l'homme

Publication date:
2001

Document Version
le PDF de l'éditeur

Contributions

Liste des intervenants

Jean-Marc Dinant

Le déploiement du nouveau plan de numérotation du réseau Internet et se risques majeurs pour la protection des données
Jean-Marc Dinant

Introduction

1. En organisant à Paris la 22ème conférence mondiale des Commissaires à la Protection des Données, la Commission Nationale Informatique et Libertés a judicieusement choisi de consacrer un atelier entier sur le thème "*les technologies pour la protection de la vie privée*". Pour sa part, cette interventio tentera de répondre à la question *: "les technologies : pour la protection de la v privée ?*". Avant de tenter d'évaluer la manière dont la technologie peut protége la vie privée, un préalable indispensable semble être d'analyser en quoi, pourquoi et comment les technologies déployées sur Internet sont "**privacides**[2]". Avant de répondre à cette question, deux remarques préliminaires s'imposent.

2. La technologie n'est pas, comme la météo, un obscur résultat de forces collossales et célestes, imprévisibles à long terme et non réglementables. La technologie Internet constitue toujours le fruit de l'interaction entre des forces sociales et finalement très humaines d'entreprises industrielles souvent multinationales, relativement prévisibles et légalement réglementables, voire reglementées.

3. L'industrie Internet produit non seulement du matériel et du logiciel mais aussi des protocoles de télécommunication. Actuellement, en Europe, la production et l'exportation ou l'importation de matériel, logiciel ou protocoles Internet ne sont pas soumises, en tant que tel, à une réglementation légale. La directive générale de 1995 vise en effet les utilisateurs de la technologie, mais non ses concepteurs ou ses vendeurs. Bizarrement, semblable réglementation légale se retrouve néanmoins aux Etats-Unis, lorsqu'il s'agit d'exporter du matériel ou du logiciel cryptographique susceptibles de protéger la vie privée des internautes[3]. Paradoxalement, la Chine a interdit l'importation des

processeurs Intel Pentium III , susceptibles de menacer la vie privée des internautes gouvernementaux[4]

4. Durant plusieurs années, l'auteur de cet article a étudié de manière direct et technique de nombreux aspects privacides de la technologie Internet. Il est impossible d'en faire un exposé détaillé dans le cadre de cette intervention. Ma quelques jalons sommaires peuvent être posés.

De quelques technologies privacides (Privacy Killing Technologies)

Les cookies

5. Les cookies sont l'objet d'une controverse depuis de nombreuses année Ce débat s'apparente à un véritable dialogue de sourds pour de nombreuses raisons. Le grand public fait rarement la différence entre :

- les cookies de session qui demeurent quelques minutes dans la mémoi vive de l'ordinateur et les cookies permanents qui sont stockés pendan plusieurs (dizaines d') années sur le disque dur;
- les cookies indiquant une caractéristique (p.e. la langue parlée) d'un internaute anonyme et les cookies contenant un identifiant global universel[5]
- les cookies issus du site visité et les cookies injectés via un hyperlien invisible par une firme de cybermarketing qui l'est tout autant

6. Outre ces trois distinctions, le phénomène des cookies ne peut être correctement appréhendé que si on le resitue dans le contexte plus large du protocole HTTP où il se développe. Quatre caractéristiques largement méconnues du protocole HTTP peuvent elles-aussi s'avérer privacides : le bavardage du programme de navigation, les hyperliens invisibles et la redirection automatique. Ces quatres éléments mis ensemble forment un cocktail explosif qui permet

- · par défaut[6]
- · à des entreprises multinationales de cybermarketing inconnues d l'utilisateur[7]
- · d'enregistrer les mots-clés tapés sur les moteurs de recherche ou les références des articles lus dans les journaux en ligne
- · en temps réel[8]
- · sur une base individuelle à l'aide d'un identificateur unique global
- · programmé pour durer jusqu'en 2035
- · plusieurs millions de fois par jour rien qu'en France

Le second clic

7. C'est le programme de navigation qui est programmé depuis de nombreuses années pour télécharger, sur le compte de l'internaute des bannières non sollicitées. Le premier clic c'est celui que l'internaute fait de

manière consciente pour obtenir de l'information d'un site en cours de visite. L second clic est effectué par le programme de navigation lui-même, vers un sit invisible, inconnu et pas nécessairement européen, au nez à la barbe de l'utilisateur et lui raconte moult détails du comportement de ce dernier.

La responsabilité des fabricants de logiciel

8. Les caractéristiques privacides du protocole HTTP 1.1 ne sont pas intrinsèques au protocole lui-même. Bien plus, durant la définition de ce protocole technique[9], les ingénieurs ont mis l'accent les dangers pour la vie privée que représentaient telle ou telle option du protocole. En effet, le mot "privacy" n'intervient pas moins de 18 fois dans ce protocole pourtant techniqu qui n'aborde pas le problème des cookies. En voici quelques extraits[10] chois :

Ø "*Having the user agent describe its capabilities in every request can be both very inefficient (given that only a small percentage of responses have multiple representations) and a potential violation of the user's privacy*" [page 6 below]

Ø " *It may be contrary to the privacy expectations of the user to send an Accept-Language header with the complete linguistic preferences of the user i every request*" [page 98]

Ø *" The client SHOULD not send the From header[11] field without the user approval, as it may conflict with the user's privacy interests or their site's securi policy. It is strongly recommended that the user be able to disable, enable, an modify the value of this field at any time prior to a request*." [page 118]

Ø "*HTTP clients are often privy to large amounts of personal information (e.g. the user's name, location, mail address, passwords, encryption keys, etc. and SHOULD be very careful to prevent unintentional leakage of this informatio via the HTTP protocol to other sources. We very strongly recommend that a convenient interface be provided for the user to control dissemination of such information, and that designers and implementers be particularly careful in this area. History shows that errors in this area are often both serious security and/ privacy problems, and often generate highly adverse publicity for the implementer's company*." [page 143]

Est-il utile de noter qu'aucun des "grands" fabriquants[12] de logiciels de navigation Internet n'a pris en ligne de compte ces recommandations ?

9. La conclusion de la page 143 rappelle la phrase devenue historique qui s trouve encore actuellement sur la page de bienvenue du site de bigbrotherinside[13], à propos du numéro de série du processeur d'Intel : "*L'avantage que ce numéro de série aurait pu nous procurer dans le domaine d la sécurité n'était pas suffisant pour contrebalancer la mauvaise réputation qu cela nous aurait donné*". Bruce Schneier, un des meilleurs cryptographes mondiaux, avit pour sa part déclaré : "

Les firmes de mauvaise réputation

10. Historiquement, c'est cette mauvaise réputation et uniquement elle qui, jusqu'à présent, a incité avec succès les producteurs de logiciels à modifier les aspects privacides. Quelques évènements historiques méritent d'être rappelés

· Au début du mois de Mars 1999, le NY Times rapportait que un identificateur spécidifque appelé identificateur global unique était systématiquement incorporé dans chaque document Word, Excel ou Powerpoint97 [14] [15]. En fait, cet identificateur global unique est basé sur le numéro de série de la carte réseau[16]. En réponse, le 8 mars 1999, Microsoft publié sur son site Web[17] deux programmes. Ceux-ci permettent, respectivement d'éviter que ce numéro soit incorporés dans de nouveaux document et d'effacer ces numéros de série des documents existants[18]. Simultanément Microsoft a annoncé que la suite bureautique Office 2000 n'insérerait plus cet identificateur global unique dans ses documents.

· C'est la pression populaire qui a (?) empêché la fusion entre les banques de données d'Abacus[19] et de Double Click. Au passage, on ne peut d'ailleur que s'étonner que la fusion entre les profils "anonymes[20]" de Double Click e la banque de données nominatives d'Abacus soit techniquement possible.

· En juillet 1999, Richard Smith un consultant en sécurité a mis en évidenc que RealJukebox, un logiciel gratuit d'écoute de CD musicaux diffusé en Europ à des millions d'exemplaires transmettait à la maison mère américaine, de manière cryptée et à intervalles réguliers les index des CDROM qui étaient insérés dans le lecteur du PC[21]. Cet événement mis en évidence l'existence de ET software[22].

Les numéros IP version 4 (Ipv4)

11. Tout réseau de télécommunication, qu'il s'agisse du réseau téléphonique o du réseau Internet, nécessite l'attribution de numéros uniques à l'échelle mondiale. Sur Internet ce numéro qui identifie une machine particulière au niveau mondial est l'adresse IP (Internet Protocol). Historiquement, ce numéro toujours été composé de quatre octet, ce qui permet d'identifier plus de quatre milliards de machines distinctes. Chaque internaute présent sur Internet peut donc être identifié par ce numéro unique.

12. Ce numéro IP peut-être attribué de manière **permanente** à une machine s celle-ci est relié à un réseau local lui-même connecté de manière permanente Internet, ou dans le cas des connexions de type DSL que ce soit via le fil téléphonique ou par le câble de télédistribution. Ce premier cas peut-être comparé à celui de l'abonné au télephone qui conserve le même numéro duran toute la durée de son abonnement.

13. Ce numéro IP peut être également attribué de façon **dynamique**, notamment lorsque l'abonné d' un fournisseur d'accès Internet se connecte au réseau Internet par le biais d'un modem téléphonique pour une session de dure

limitée (quelques minutes, voire plusieurs heures). Lors des sessions ultérieures, le même internaute se verra attribuer par le même fournisseur d'accès un numéro IP différent du précédent. Cette situation peut-être comparé à celle d'une personne ne disposant pas d'un abonnement téléphonique et n'utilisant que des cabines téléphoniques et rarement les mêmes.

14. Toutefois, contrairement au réseau téléphonique[23]*, il n'y a pas, sur le réseau Internet de moyens simples de masquer le numéro d'appel de la machine appelante*. La seule solution actuelle consiste à passer par un tiers d confiance (anonymiseur) qui masquera cette adresse IP au réseau en y substituant la sienne. Les procédés d'anonymisation restent toutefois globalement peu fiables, ralentissent le fonctionnement du réseau et nécessite en général un paiement.

Les numéros IP version 6 (Ipv6)

15. Pour des raisons techniques, il semble que cet espace de quatre milliards d'adresses soit proche de la saturation. La structure de cette addresse IP est donc en train de se modifier et sera notamment portée de 4 octets à 16 octets Dans ces 16 octets, le protocole Ipv6 recommande que 6 de ces 16 octets soient constitués par le numéro de série électronique de la carte réseau Ethernet présente sur l'ordinateur personnel. Ce numéro de série (addresse MAC (Medium Access Control)) est un numéro unique au monde gravé dans l'électronique de la carte réseau de type Ethernet, le standard pour constituer des réseaux locaux connectés à Internet. Ce numéro de série posait un problème de vie privée auparavant mais restait normalement interne au réseau local et n'était pas transmis sur le réseau Internet.

16. Avec le déploiement de la nouvelle numérotation Ipv6, chaque machine e donc chaque internaute, transmettra, le plus souvent à son insu, et sans qu'il puisse s'y opposer, un numéro de série unique au monde et stable dans le temps. Ce numéro sera transmis quelque soit le service utilisé sur Internet : envoi de courrier électronique, forums de discussion, accès aux moteurs de recherche (pour chercher les horaires des offices des mosquées, l'adresse d'u syndicat, du viagra ou des traitements pour guérir le sida, pour lire certains articles des journaux en ligne, etc.

17. Avec Ipv6, il n'y a plus de différence, au regard de la protection des données, entre l'attribution statique ou dynamique, pour la simple raison que l numéro de série MAC sera toujours une partie de l'adresse IP, quelque soit le fournisseur d'accès Internet

18. Le tableau ci-dessous montre l'importance de ce numéro par rapport à quelques risques plus classiques qui ont déjà été identifiés. Tous ces risques concernent la transmission sur le réseau d'un numéro unique au monde[24]. C tableau décrit la situation par défaut de "l'internaute de la rue". Il est clair que le "riches" et les "malins" qui en savent d'avantage peuvent se protéger bien mieux.

Tableau 1. Privacy scoring de quelques technologies privacides

Nom	*% intern. concernés[25]*	*Durée de vie moyenne*	*Infor-mation*	*Oppo-sition*	*Transmission à tout tiers*	*Transmissio à un tiers*
Processor Number	40¯	Celle de l'ordinateur	Oui	Oui	Non	Oui
Microsoft GUID	40¯	Celle du document	Non	Non-Oui	Non	Oui
Cookie de session	100®	+/- 20 minutes	Non	Non	Non	Oui
Ipv4 dynamique	50¯	Minutes ou heures	Non	Non	Oui	Oui
Cookie permanent	98¯	Celle de l'ordinateur	Non	Non	Non	Oui[26]
Ipv4 permanent	50¯	Qqs mois ou années	Non	Non	Oui	Oui
MAC Address	90	Celle de la carte réseau	Non	Non	Non	Non
Ipv6 permanent		Celle de l'ordinateur	Non	Non	Oui	Oui
Ipv6 dynamique		Celle de l'ordinateur	Non	Non	Oui	Oui

Vert : Pas ou peu privacide ou en cours de résolution.

Rouge : Privacide. A éviter si possible.

Noir : Hautement privacide. A éviter à tout prix.

Les leçons non apprises de l'histoire

19. Chacun gardera en mémoire l'histoire du PSN d'Intel, qui a d'ailleurs rencontré le Groupe 29 à ce sujet en 1998[27]. Sous la pression populaire (http://www.bigbrotherinside.com)/, Intel, un géant de l'industrie du hardware Internet, a du faire marche arrière et a supprimé ce numéro de série en 2000. Microsoft a du faire la même marche arrière lors de la découverte de l'incorporation de ce fameux numéro MAC dans chaque document Word, Exce ou Powerpoint 97.

20. Dans le cadre d'IPv6, l'enjeu est sans commune mesure avec le PSN d'int ou le GUID de Microsoft. Il ne s'agit plus de doter une version d'un processeu d'une marque particulière ou certains documents d'un numéro de série qui, dar certains cas et toujours avec l'accord de la personne concernée peut être transmis sur le réseau. Il s'agit présentement de d'incorporer de manière systématique dans toutes les communications Internet un numéro de série présent sur la grande majorité des ordinateurs personnels, sans que la personn

concernée n'en soit informée, ni, a fortiori, ne puisse s'y opposer.

Conclusions

L'interdiction des identifiants globaux uniques (GUID)

21. Au regard de la protection des données, l'utilisation des identifiants globau universels (GUID) doit être systématiquement interdite. Elle contrevient de manière évidente aux principes élementaires de sécurité contenus dans les articles 16 et 17 de la directive générale 95/46. Si deux traitements poursuiver des finalités différentes voire incompatibles, il doit être rendu aussi techniquement difficile que possible d'effectuer un rapprochement entre les données d'un individu inscrit dans ces deux traitements. Un mesure **élémentaire** de sécurité propre à prévenir ce rapprochement non autorisé est d doter un même individu d'identifiants différents selon le traitement auquel il participe. Ceci est spécialement vrai lorsqu'un ou plusieurs des traitements en question poursuit une finalité sensible, judiciaire ou médicale. Le responsable d traitement qui manque à cette obligation élémentaire de sécurité contrevient au articles 16 et 17 de la directive. Dans les cas ou des rapprochements ponctuel doivent être possible au cas par cas, il convient que cet identifiant unique soit encrypté[28] à l'aide d'une clé secrète propre à chaque responsable de traitement et à chaque traitement.

Le contrôle des logiciels et matériels

22. Il est difficile de baser ce contrôle sur la directive générale 95/46 parce qu les concepteurs de logiciel ou de matériel ne gèrent pas directement les données. Cette considération est aujourd'hui empreinte d'une très grande naïveté technologique. Au fil de la dernière décennie, l'utilisateur des technologies informatiques en général et de produits liés en particulier se trouv face à une immense usine à gaz effectuant des milliards d'opérations par seconde et dont la logique le dépasse. L'écran n'est plus qu'un pâle reflet des données qui circulent sur le réseau. De nombreux traitement invisibles sont effectués par dessus son épaule.

La directive 1999/5/CE du Parlement européen et du Conseil, du 9 mars 1999 concernant les équipements hertziens et les équipements terminaux de télécommunications et la reconnaissance mutuelle de leur conformité définit comme (art. 2 (b)) "équipement terminal de télécommunications", un produit permettant la communication, ou un composant pertinent d'un produit, destiné être connecté directement ou indirectement par un quelconque moyen à des interfaces de réseaux publics de télécommunications. Un simple programme d navigation ou de courrier électronique ou encore un routeur peuvent donc être considérés comme équipements terminaux de télécommunication. Dans son article 3 c (exigences essentielles), la même directive pose, que la Commissio peut décider *que les appareils relevant de certaines catégories d'équipements ou certains types d'appareils sont construits de sorte qu'ils comportent des sauvegardes afin d'assurer la protection des données à caractère personnel e de la vie privée des utilisateurs et des abonnés*. La commission Européenne

possède donc là un instrument juridique contraignant et directement disponible

La concurrence loyale dans le marché des programmes Internet

23. Un programme de navigation est nettement plus complexe à réaliser qu'u logiciel serveur Web. Or le prix de ces deux types de programmes est inversément proportionnel à leur coût. Les premiers sont gratuits et les deuxièmes sont de l'ordre du millier d'euros. C'est une vue de l'esprit. En fait, l société qui achète un serveur web paie implicitement pour que les logiciels clients permettant d'accèder à son serveur soient distribués de manière gratuit Cette situation entraîne deux biais à terme préjudiciable pour la protection des données de l'Internaute.

24. Les clients qui rapportent de l'argent à l'industrie du logiciel Internet sont le entreprises qui, de manière naturelle, désirent collecter le maximum d'informations sur leurs visiteurs virtuels. Les Internautes représentent un coû pour cette entreprise et la protection de leur données pourrait être considérée une moins value pour les produits qu'elle distribue. Entre un serveur qui respecte l'anonymat des visiteurs du site ou un serveur permettant une connaissance intime, automatique et personnalisée de chaque visiteur, leque une entreprise de commerce électronique aura-t-elle tendance à pencher ?

25. Les entreprises qui désirent se lancer dans la production de programmes d navigation payants[29] subissent une concurrence déloyale de la part de celle qui distribuent ces mêmes produits de manière gratuite.

---

[1] Jean-Marc Dinant (http://www.droit.fundp.ac.be/cv/jmdinant/) est également directeur de recherches au Centre de Recherche Informatique et Droit de l'Université de Namur

[2] A comprendre comme insecticide ou paricide : tueur de vie privée. L'aute ose ce néologisme afin de ne rien perdre de la force du mot anglais "privacy killing".

[3] "*NSA provides the Department of State with technical advice to determin whether the commodity is a cryptographic system, equipment, assembly, module, integrated circuit, component or software "with the capability of maintaining secrecy or confidentiality of information*" covered under Category XIII(b)(1) of the United States Munitions List ("USML"). 22 C.F.R. § 121.1, XIII(b)(1). Cfr http://people.qualcomm.com/karn/export/crowell.html, visité en août 2001.

[4] Guangming Daily newspaper, Wednesday, June 30, 1999 disponible su http://jya.com/cn-p3-peril.htm. dernière visite en août 2001.

[5] Global Unique Identifier (GUID). Tel le cookie attribué par Double Click

[6] Les ingénieurs et utilisateurs avertis pourront facilement trouver sur le réseau des milliers de programmes divers permettant, dans le meilleur des cas de se protéger gratuitement. Dans le monde réel, il est aussi possible de porte un gilet pare balles et/ou un masque à gaz et de rouler dans une voiture blindé flanquée de quelques garde du corps. Privacy for the clever and for the rich ?

[7] L'exemple type est bien évidemment Double Click qui produisait en 200 plus d'un demi-milliard de bannières publicitaires sur le réseau chaque mois.

[8] Techniquement, l'entreprise de cybermarketing connaît le profil de l'internaute AVANT de transmettre la bannière

[9]

[10] The page numbering indicated between brackets refer to the numeratio of W3C.

[11] Note of the author : From header field is used for naming the referring page

[12] Opéra version 3 et 4 permet à l'internaute de bloquer l'envoi de la page référente

[13] http://www.bigbrotherinside.org/

[14] http://www.junkbusters.com/ht/en/microsoft.html#history

[15] http://www.techserver.com/noframes/story/0,2294,25591-41382-304399 0,00.html

[16] Plus précisément sur l'adresse MAC (Medium Access Control) de la cart Ethernet

[17]

[18] Notons au passage que l'utilisation de ces outils nécessite dans certains cas une mise-à-jour de Windows 95, en téléchargeant plus de trente millions d'octets sur le site de Microsoft

[19] a *cooperative membership database, contains records from more than 1,100 merchandise catalogs, with more than 2* ***billion*** *consumer transactions from virtually all U.S. consumer catalog buying households* http://www.abacus direct.com

[20] http://www.doubleclick.net/company_info/about_doubleclick/privacy . : *DoubleClick does not collect any personally-identifiable information about you such as your name, address, phone number or email address.*

[21] http://www.thatworld.com/news/realjukebox.html

[22] L'image, extraite du film de *Steven Spielberg* est parlante. L'extraterrestr ET "téléphone" en cachette de temps en temps à la maison, pour raconter ce q s'est passé sur la terre. Un *ET software* est donc un programme qui communique *via* Internet des détails sur le comportement de son utilisateur. U exemple célèbre est le cas de *RealJukeBox Player*, *software* d'écoute de CD musicaux diffusé à plus de treize millions d'exemplaires qui rapportait régulièrement à la société mère (*RealNetworks*) le détails de CD insérées dan le lecteur, de manière encryptée, <http://www.tiac.net/users/smiths/privacy/realjb.htm>. À la suite d'un article par dans le *New York Times*, *RealNetworks* a modifié son logiciel.

[23] Au niveau technique, ce numéro d'appel est TOUJOURS transmis sur le réseaux numérisés. Si le numéro est censé être secret, un bit particulier indiquera qu'il ne pourra être révélé (SIC).

[24] Durant sa durée de vie, numéro unique identifiant un ordinateur par rapport à un tiers.

[25] Cette estimation a été faite ex æquo et bono.

[26] Il est possible de partager un cookie entre plusieurs ordinateurs serveur du même sous domaine.

[27] C'est dans ce contexte que le Groupe 29 a produit la recommandation su les traitements invisibles effectués par hardware ou par software (http://europa.eu.int/comm/internal_market/en/dataprot/wpdocs/wp17fr.pdf)

[28] "compte tenu de l'état de l'art", comme précisé dans l'article 17 de la Directive 95/46. Dans ce cadre, il est opportun de considérer le hacking comm un "art".

[29] par exemple Opera : http://www.operasoftware.com

Liste des intervenants

The arrival of the new Internet network numbering system and its major risks to data protection

By **Jean Marc Dinant**[1]

Introduction

1. By organising the 22nd world conference of Data Protection Commissioners in Paris, the Commission Nationale Informatique et Libertés quite intelligently chose to dedicate a whole workshop to the subject of "Privacy Enhanced Technologies". The purpose of this paper is to answer the question : *Technologies : Enhancing Privacy ?"*. Before attempting to assess how technology can protect privacy, a prerequisite seems to be the analysis of what, why and how technologies used on Internet are **"**privacy killing**"**[2]. Before trying to answer that question, two comments need to be made.
2. Technology is not, like weather forecasting, the obscure result of tremendous and celestial forces, unpredictable on the long run and uncontrollable. The Internet technology is the fruit of interaction between social and very humanly forces of industrial concerns, often international, relatively predictable and legally controllable, and sometimes controlled.
3. Not only does the Internet industry generate equipment and software programmes, it also produces telecommunication protocols. Currently in Europe, the production and exportation or importation of hardware, software or Internet protocols is not subjected, as such, to any legal regulation. The general directive of 1995 is aimed at technology users, not at technology designers or sellers. Strangely, similar legal regulation is found in the United States, when exporting cryptographic hardware or software likely to protect the privacy of Internet surfers[3]. Quite paradoxically, China prohibited the importation of Intel Pentium III processors, which were likely to threaten the privacy of Internet surfing government employees[4].
4. For several years, the author of this paper carried out a direct and technical study of numerous privacy-killing aspects of the Internet technology. It is impossible to give a detailed account of the studies in this paper. However, a few brief indications can be given.

A few privacy killing technologies

Cookies

5. Cookies have been controversial for many years. The debate is a dialogue of the deaf, for many reasons. Rarely does the general public know the difference between :

---

1 Jean Marc Dinant (http://www.droit/fundo.ac.bc/cv/jmdinant) is computer expert to the Belgian Data Protection Commission and also head of research at the Centre de Recherche Informatique et Droit de l'Université de Namur. This text has been written with the support of ECLIP (http://www.eclip.org), a european project funded by the EC.

2 "Privacide" in original french text

3 "NSA provides the Department of State with technical advice to determine whether the commodity is a cryptographic system, equipment, assembly, module, integrated circuit, component or software" with the capability of maintaining secrecy or confidentiality of information" covered under Category XIII (b) (1) of the United States Munitions List ("USMA"), 22 CFR § 121.1 XII (b) (1) Cfr http://peoplequalcomm.com/karn/export/crowell.html, visited in August 2001.

4 Guangrning Daily Newspaper, Wednesday, June 20, 1999, available at http://jva.com/cn-p3-peril/http, last visit in August 2001

- session cookies, which remain in the computer's memory for a few minutes, and permanent cookies, which are stored on the hard disk for several tens of years
- cookies showing a characteristic (for instance, the language) of a given internet surfer, and cookies containing a Global Unique Identifier[5]
- cookies from the visited site and cookies injected via an invisible hyperlink by an also invisible cybermarketing firm.

6. In addition to these three distinctions, the cookie phenomenon cannot be understood if it is not placed in the broader context of the HTTP protocol where it develops. Four widely unknown features of the HTTP protocol can also prove privacy killing : the chattering of navigation programmes, invisible hyperlinks, and automatic redirecting. Put together, these four elements, form an explosive mixture enabling
   - by default[6]
   - international cybermarketing companies unknown to the user[7]
   - to use the key words types on search engines or references of papers read in on-line newspapers
   - in real time[8]
   - on an individual basis, using a global unique identifier
   - programmed to last until 2035
   - Several million times a day, in France alone.

The second click

7. The navigation program itself, programmed for several years to download to the Internet user account unsolicited banners performs this second click''. The first click is the click the Internet user does consciously in order to obtain information during a visit. The second click is done by the navigation programme, towards an invisible site, unknown and not necessarily European, unknown to the user, telling many details about the user's behaviour.

The software manufacturer's responsibility

8. The privacy killing characteristics of the HTTP 1.1. protocol are not inherent to the protocol itself. Furthermore, when defining the technical protocol[9], the engineers that designed it emphasised the dangers to privacy such protocol option constituted. Indeed, the word "privacy" comes up no less than 18 times in this otherwise technical protocol, which does not discuss the cookie problem. Below are a few selected extracts[10].
   - *"Having the user agent describe its capabilities in every request can be both very inefficient (given that only a small percentage of responses have multiple representations) and a potential violations of the user's privacy (page 68 below).*
   - *It may be contrary to the privacy expectations of the user to send an Accept Language header with the complete linguistic preferences of the user in every request (page 68).*
   - *The client SHOULD not send the From header [11] field without the user's approval, as it may conflict with the user's privacy interests or their site's security policy. It is*

---

5 Global Unique Identifier (GUID). Such as the cookie assigned by Double Click

6 Knowledgeable engineers and users can easily find thousands of different programmes allowing to protect oneself, possibly free of charge. In the real world, one can also wear a bullet-proof jacket and/or a gas mask and travel in an armoured vehicle guarded by a few bodyguards. Privacy for the clever and for the rich?

7 One typical example is Double Click, which produced over half a billion advertising banners on the network in 2000.

8 Technically, the cybermarketing company knows the internet user's profile BEFORE sending the banner.

9 Hypertext Transfer Protocol -- HTTP/1.1 http://www.w3.org/Protocols/rfc2068/rfc2068

10 The page numbering indicated between brackets refer to the numeration of W3C

11 Note of the author: From header field is used for naming the referring page[18].

strongly recommended that the user be able to disable, enable and modify the value of this field at any time prior to a request" (page 118).

- *HTTP clients are often privy to large amounts of personal information (e.g. the user's name, location, mail address, passwords, encryption keys, etc.), and SHOULD be very careful to prevent unintentional leakage of this information via the HTTP protocol to other sources. We very strongly recommend that a convenient interface be provided for the user to control dissemination of such information, and that the designers and implementers be particularly careful in this area. history shows that errors in this area are often both serious security and/or privacy problems, and often generate highly adverse publicity for the implementer's company". (page 143).*

Must it be added that none of the major manufacturer[12,13] of Internet navigation software programmes took the above recommendations into account ?

9. The conclusion of page 143 reminds the now historical phrase that is still currently on the welcome page of the bigbrotherinside site[14], about the serial number of the Intel processor :*"the advantage this serial number could have given us in the area of safety was not enough to offset the bad reputation it would have generated"*. Bruce Schneier, one of the top cryptographs in the world, said[15]: *" As a cryptographer, I cannot design a secure system to validate identification, enforce copy protection, or secure e-commerce using a processor ID. It doesn't help. It's just too easy to hack."*

Companies with a bad reputation

10. Historically, bad reputation, and only it, successfully prompted software manufacturers to change privacy-killing aspects. Here are a few historical events :
    - in early March 1999, the NY Times reported that a specific identifier called Global Unique Identifier was systematically introduced in each Word, Excel or Powerpoint 97 document[16, 17]. Actually, the global unique identifier is based on the serial number of the network card[18]. In answer, on March 8, 1999, Microsoft published two programmes on its Web site[19]. They helped prevent the number from being introduced in new documents and erase such serial numbers from existing documents[20]. Simultaneously, Microsoft announced that the Office 2000 sequel would no longer introduce the global unique identifier in its documents.
    - High pressure from users prevented (?) the merging of the Abacus[21] and Double Click databases. By the way, one can only wonder that the merging be technically possible between the "anonymous[22]" profiles of Double Click and Abacus' nominative database.

---

12 Opera version 3 and 4 enable the internet user to block the sending of the referring page

13 Mozilla permit to stop automatic downloading of images not located on the visited web site (i.e. ad bannering performed by an invisible third party)

14 http://www.bigbrotherinside.org

15 Cited in http://www.zdnet.com/zdnn/stories/comment/0,5859,2194863,00.html

16 http://www.junkbusters.com/ht/cn/microsoft.html/history

17 http://www.techserver.com/noframes/story/0,2294,25591-41382-304399-0,00.html

18 Specifically on the MAC (Medium Access Control) address of the Ethernet card

19 http://www.microsoft.com/PressPass/features/1999/03-08custletter2.asp

20 Note that the use of these tools requires in some cases an updated of Window 95, by uploading over thirty million octets from Microsoft's site.

21 a *cooperative membership database, contains records from more than 1,100 merchandise catalogs, with more than 2* ***billion*** *consumer transactions from virtually all U.S. consumer catalog buying households*" http://www.abacus-direct.com

22 http://www.doubleclick.net/company_info/about_doubleclick/privacy : *" DoubleClick does not collect any personally-identifiable information about you, such as your name, address, phone number or email address.*

- In July 1999, safety consultant Richard Smith showed that RealJukeBox, a free software programme for listening to music CDs broadcast in Europe in millions of copies, sent to the parent company in the United States, on a regular basis, encrypted index information found on the CDROMS inserted in the PC's player[23]. The event underlined the existence ET software[24].

IP version 4 numbers

11. Any telecommunication network, whether it be a phone network or the Internet network, requires the assignment of unique numbers on a world scale. The Internet number identifying a particular machine in the world is the IP address (Internet Protocol). Historically, the number has always had four octets, which allows the identification of over four billion specific machines. Each Internet user can be identified by this unique number
12. The IP number can be assigned in a permanent manner to a machine connected to a local network itself connected to Internet permanently, or in the case of DSL type connections whether it be via a telephone line or through a teledistribution wire. The first case can be compared to a telephone user keeping the same number for the whole time of his/her subscription.
13. The IP number can also be assigned in a dynamic way, including when a client of an Internet access provider connects to the Internet via a telephone modem for a session limited in time (a few minutes or even several hours). During later sessions, the same Internet user will be assigned by the same access provider a different IP number. This situation can be compared to that of an individual who does not have a telephone service and only uses telephone booths, and rarely the same one.
14. However, contrary to the telephone network[25], *there is not any simple way to conceal the call number of the calling machine on the Internet network.* The only current solution is to go through a trusted third party (anonymiser) that will hide the IP address from the network, by substituting its own address. However, anonymisation processes are not reliable, slow down the network and require payment.

IP version 6 numbers (Ipv6)

15. Due to technical reasons, it seems that the four billion-address space is close to being saturated. The structure of the IP address is changing, and will be increased from 4 to 16 octets. Of these 16 octets, the Ipv6 protocol recommends that 6 octets be formed by the electronic serial number of the Ethernet network card of the personal computer. This serial number (MAC address – Medium Access Control) is a unique number engraved in the electronic system of the Ethernet type network card, the standard for constituting local networks connected to Internet. This serial number used to pose a privacy problem but remained an internal local network, and was not normally sent out on the Internet network.
16. With the implementation of the new Ipv6 numbering system, each machine, and each Internet user, will be sending out, often without their knowledge, and without any way to

---

[23] http://www.thatworld.com/news/realjukebox.html

[24] The image is from Steven Spielberg's movie and is meaningful. ET alien phones home from time to time, without anyone knowing, to tell what is happening on earth. An ET software is a programme that communicates, via the Internet, details on the user's behaviour. One famous example is the RealJukeBox Player, a music CD listening software programme broadcast in more than thirty million copies, providing the parent company (RealNetworks) details about the CD inserted in the player, in an encrypted way, http://www/tiac/net/users/smiths/privacy/realjb.htm. After a paper in the New York Times, Real Networks changed its software.

[25] Technically, this calling number is ALWAYS send on digital networks. If the number is supposed to be secret, a particular bit will indicate that it cannot be disclosed

oppose, a serial number unique in the world and stable in time. This number will be sent, whatever the Internet service used : e.mail sending, chat forums, search engine access (to look for mosque service hours, the address of a union, viagra or AIDS-curing treatments, to read some on-line newspapers, etc).

17. With Ipv6, there is no difference, in terms of data protection, between static and dynamic assignment, simply because the MAC serial number will always be part of the IP address, whatever the Internet access provider.
18. The following table shows the importance of the number in relationship to a few more ordinary risks that have already been identified; all these risks pertain to the transmission of a number unique in the world[26] onto the network. The table describes the default situation of the "average Internet user". Clearly, the rich and the clever, which are more knowledgeable, can have better protection.

Table 1 : Privacy scoring of a few privacy killing technologies

Nom	*% concerned netizens[27] - trends*	*Mean Lifetime*	*Infor-mation*	*Oppo-sition*	*Transmission to any third party*	*Transmission to one single third party*
Processor Number	40%⟶	**Computer Lifetime**	Yes	Yes	No	Yes
Microsoft GUID	40%⟶	Document LifeTime	No	No-Yes	No	Yes
Session cookie	100%→	+/- 20 minutes	No	No	No	Yes
Dynamic Ipv4	50%↓	Minutes or Hours	No	No	**Yes**	Yes
Permanent Cookie	98%↓	**Computer Lifetime**	No	No	No	Yes[28]
Permanent Ipv4	50%↓	Many Months or Years	No	No	**Yes**	Yes
MAC Address	90%↑	**Computer Lifetime**	No	No	No	No
Permanent Ipv6	↑↑	**Computer Lifetime**	No	No	**Yes**	Yes
Dynamic Ipv6	↑↑	**Computer Lifetime**	No	No	**Yes**	Yes

Green	: not or not much privacy killing. Or under way to be solved.
Red	: Privacy killing. To be avoided whenever possible.
Black	**: Higly privacy killing. To be avoided at any price.**

The unlearnt lessons of history

19. Everyone will remember the story of the PSN of Intel regarding this matter[29]. Under pressure from users (http://www.bigbrotherinside.com), Intel, an Internet hardware industry giant, had to back down and removed the serial number in 2000. Microsoft had to back down too, when it was discovered that the famous MAC number was being introduced ion each Word, Excel or PowerPoint 97 document.
20. In the framework of IPv6, the stakes have nothing to do with Intel's PSN or Microsoft's GUID. This is no longer about giving a processor version a particular mark or documents a serial number that in some case and possibly with the consent of the user can be sent out to the network. It is about introducing, systematically, in all Internet communications, a serial number incorporated in most personal computers, without the user knowing, let alone being able to do anything about it.

---

26 During its lifetime, unique number identifying a computer from a third party

27 . This estimation is made ex æquo et bono

28 It is possible to share a cookie between several server computers of the same sub-domain.

29 It is then that Group 29 produced its recommendation invisible processing carried out by hardware or by software (http://europa.eu.int/comm/internal_market/en/dataprot/wpdocs/wp17fr.pdf)

Conclusions

Prohibition global unique identifiers (GUID)

21. With respect to data protection, the use of global unique identifiers (GUID) should be strictly prohibited. It is an obvious violation of basic safety principles as contained in articles 16 and 17 of general directive 95/46. If two processes pursue different or even incompatible objectives, it should be as technically difficult as possible to bring together the data of an individual registered in both processes. A basic safety measure helping prevent such unauthorised bringing together consists in giving one individual different identifiers, depending on the process involved. This is especially true when one or several of the processes pursue a sensitive, legal or medical objective. Data controllers violating such a basic safety duty violate articles 16 and 17 of the directive. In the event of occasional reconciliation being possible on a case by case basis, the unique identifier should be encrypted[30] using a secret key specific to each process manager and each process. By using this tip, a cross profiling can only be performed with the consent of the involved data controllers.

From a static to a dynamic Mac Address

22. For many years, the identifying number of the Ethernet Card has been not much privacy killing, just because this serial number was remaining local in the context of an intranet and was not relayed beyond the router, classic gateway between a Local Area Network and the whole Internet. But, however, the single presence of this global unique identifier has yet caused a clash since he has been invisibly incorporated by Microsoft in all Word, Excel and PowerPoint 97 documents. Nothing, at the technical level, justifies a global unique identifier. Technically speaking, it is widely sufficient that each network card in the same LAN has a different Id to avoid technical problems. Typically, a Local Area Network will count many tens or hundreds machines. It is possible that this number can be randomly generated or choose by the user it self[31].

Hardware and software control

23. It is difficult to found such control on general directive 95/46 because software and hardware designers do not manage the data directly[32]. This consideration now appears to be very naive[33]. During the course of the last decade, the user of computer technologies in general and of linked products in particular faced an enormous facility making billions of transactions per second, with confusing methods. The screen is only a small reflection of the data circulating on the network. Numerous invisible processes are being performed over the user's shoulder.

    The 1999/5/CE directive of the European Parliament and Council, dated March 9, 1999, relating to wireless networks and terminal telecommunication equipment and their mutual acknowledgement of compliance defines as (art 2 b): "a telecommunication terminal", any product allowing communication, or any component of a product intended to be directly or indirectly connected by any means to public telecommunications network interfaces.

---

[30] Considering the state of the art" as indicated in article 17 of Directive 95/46. To this end, it is advisable to consider hacking as an "art".

[31] This possibility is explicitly foreseen in RFC 3041Privacy Extensions for Stateless Address Autoconfiguration in IPv6 (http://www.ietf.org/rfc/rfc3041.txt?number=3041)

[32] This remark does not concern privacy killing privacy killing systems like Passport or Hailstorm desinged and managed by Microsoft and that seems to send back to Microsoft many details of the behaviour of Windows XP.(cfr EPIC complaint : http://www.epic.org/privacy/consumer/MS_complaint.pdf)

[33] Directive 95/46 first drafts has been issued when the Internet was marginal and when Windows 95 was not yet distributed on the majority of personal computers.

Any simple navigation or e-mail programme, or even a router can be considered as terminal telecommunication equipment. In its article 3 c (main requirements), the same directive provides that the Commission can decide that *machines in some categories of equipment or some types of machines should be built so as to include safeguards to protect personal and private life information of the users and subscribers.* This is a legal and binding instrument available to the European Commission.

Unfair competition in the market of Internet programmes.

24. A navigation programme is significantly more complex to build than a Web server software programme. However, the price of both types of programmes is inversely proportional to their cost. The first are free, and the latter cost approximately one thousand Euro. That is a purely theoretical view. Actually, a company buying a web server pays for the client software providing access to its server being freely distributed. This situation has tow consequences, which are harmful to user data protection.
25. Clients bringing money to the Internet software industry are companies who, quite naturally, want to collect as much information as possible about their virtual visitors. Internet users represent a cost to these companies, and protecting their data can be considered as a cost of the products they distribute. Between a server respecting the anonymity of site visitors, and a server providing in-depth, automatic and personalised knowledge of each visitor, which will an electronic trading company tend to choose?
26. Companies wanting to engage in the production of paying navigation programmes[34] are subjected to unfair competition from those distributing those products free of charge.

[34] . for instance Opera : http://www.operasoftware.com